

## Speciální statistické postupy při práci s výběrovými daty získanými z malých populací

Cílem předloženého textu je položit teoretické a praktické základy pro korektní statistickou práci s daty, která byla získána z malých populací, pokud z této populace vybíráme nezanedbatelnou část (nad 5 %). Využíván je koncept konečnostního násobitele, který je běžně (ovšem bez dalšího) uváděn i v základních učebnicích statistiky. Na příkladech *t*-testů (jedno- a dvouvýběrového), analýzy rozptylu, korelace, lineární regrese a kontingenčních tabulek je ukázáno využití násobitele pro vyhodnocení statistického testu, resp. pro interval spolehlivosti. Pro maximální praktičnost kromě vzorců a příkladů byla připravena též pomůcka v excelu, která umožňuje poměrně snadné výpočty. Cílem článku je poukázat na problematiku statistického zpracování výběrů ze základních souborů, které jsou malé (do řádu tisíců). Kromě ukázek statistického zpracování základními postupy inferenční statistiky (*t*-test, analýza rozptylu, korelace, regresní analýza a chí-kvadrát test pro proporce či kontingenční tabulky) bude na počátku věnována též pozornost realizaci výběrů.

Ještě před hlubším vnořením se do tématu je možné poznamenat, že pokud provádíme výběry z malých populací a tyto výběry tvoří nezanedbatelnou část populace (v literatuře se běžně uvádí nad 5 či 10 %), je vhodné užít dále představené postupy. Na jednu stranu je to sice těžší (nejsou běžně implementovány ve statistickém softwaru a je třeba používat ruční výpočty či speciální pomůcky). Tuto nesnáž ale kompenzuje skutečnost, že v uvedených případech by bez použití speciálních postupů často výzkumník skončil konstatováním o nemožnosti zamítnutí nulové hypotézy, které by ovšem bylo chybné s ohledem na nesprávný výpočetní postup. Využití speciálních postupů také umožňuje stanovit lépe odhady požadované velikosti výběrových souborů, které mohou být *ceteris paribus* u výběrů z malých populací menší než při vybírání z velkých, resp. z teoreticky nekonečných populací. Článek doprovází pomůcka, která umožňuje snadno realizovat popsané výpočty.<sup>1</sup>

Volně navazují na text Soukupa a Kočvarové (2016), proto zde nejsou opakovány ani definovány základní pojmy a postupy, tj. populace, výběr, reprezentativita a metody vybírání. Je nutné opakovat pouze jediný poznatek z odkazovaného textu, a to skutečnost, že postupy popsané v tomto textu jsou uplatnitelné jen tehdy, pokud je výběr prováděn náhodně (pravděpodobnostně), pro nenáhodné výběry dále uváděné

<sup>1</sup> Návodné pomůcky jsou vytvořeny v prostředí MS Excel, se kterým běžně výzkumníci umějí zacházet, a není tak třeba dalšího výkladu. Obdobné pomůcky lze nalézt ve formě on-line kalkulaček na internetu.

**126** postupy a vzorce neplatí, a nelze je tedy bez dalšího používat. Namísto je zároveň upozornit, že problematika se jistě nevyčerpává zmíněnými základními statistickými postupy, nicméně s ohledem na možnosti rozsahu jednoho textu se autoři omezili pouze na základní postupy, které jsou nejčastěji užívány českými pedagogickými výzkumníky pro zpracování kvantitativních studií.

Před dalším rozbohem je ještě třeba předejít jednomu potenciálnímu nedorozumění. Pokud se hovoří o výběrech z malých populací, pak to nutně neznamená, že výběr musí být malý. Pro úplnost dodejme, že o malých výběrech hovoří statistika v případech, kdy velikost výběrového souboru nepřesahuje cca 30–50 jednotek. To, že je výběr malý, pak může mít dopady na statistické zpracování (např. v případě malého výběru je zapotřebí dodržet striktně požadavek na normální rozdělení zkoumané veličiny). Pokud hovoříme o výběrech z malých populací (v řádu stovek či maximálně jednotek tisíců), pak výběr může být velký. S ohledem na dále popsané platí, že čím větší proporce z populace vybíráme, tím více je potřebné používat dále uvedené postupy.

## 1 Příklady malých populací v pedagogickém výzkumu

Běžně zkoumané obecné populace v pedagogickém výzkumu rozhodně mezi malé nepatří. Pokud se zaměřujeme na žáky konkrétního ročníku základní školy, pak jich máme v České republice cca 100 tisíc, zaměříme-li pozornost na učitele základních či středních škol, hovoříme o řádech desítek tisíc. Nicméně pokud naše pozornost cílí na určitý specifický region či specifické oblasti, může být situace už zcela jiná. Budeme-li chtít například zkoumat žáky praktických škol, máme k dispozici již „jen“ 18 tisíc žáků, v jednotlivém ročníku už budeme na úrovni jednotky tisíců. A obdobná situace nastane při jakémkoli specifičtějším vymezení cílové populace. Samozřejmě ideálním postupem pro výzkumnou práci s malou populací by bylo zjistit informace od všech jejích jednotek. To ale samozřejmě prakticky narazí na nedostatek financí a času, proto provádíme i z těchto populací výběry. Výsledné výběry tak mohou být v řádech desítek či stovek jednotek, velikost výběru by měla vždy odpovídat výzkumným záměrům a zejména by měla odrážet požadavky na dostatečnou velikost s ohledem na plánované statistické zpracování – podrobnosti opět viz text Soukupa a Kočvarové (2016). Je namísto zopakovat, že v případě, kdy provádíme výběr z malé populace a chceme výsledky na tuto populaci zobecňovat statistickými postupy, musíme dodržet striktně pravidla náhodného (pravděpodobnostního) vybírání. Jen tak lze alespoň částečně zabránit nežádoucímu efektu, kdy nesprávně vybraná jednotka (s extrémními charakteristikami) může výrazně ovlivnit naše analytické výsledky. V této fázi můžeme narazit na dva problémy. Prvním může být skutečnost, že nemáme k dispozici oporu výběru, tj. seznam všech jednotek z námi zkoumané populace, a nelze tedy snadno provést prostý náhodný výběr. Situace je řešitelná složitějšími designy vícestupňových či skupinkových výběrů, nicméně tato řešení mají své důsledky popsané v textu Soukupa a Kočvarové (2016). Druhým problémem bývá nulová

či minimální informace o struktuře námi zkoumané populace, díky čemuž nemáme možnost ověřit reprezentativitu námi získaného výběru. Ideální situace samozřejmě nastává, pokud máme kompletní (a pokud možno aktuální) oporu a pokud navíc známe strukturu naší populace.

## 2 První poznámky o velikosti výběru z malých populací (konečnostní násobitel)

Jak již bylo náznakem uvedeno, za určitých podmínek platí, že pokud vybíráme z malých populací, potřebujeme pro stejně přesné výsledky ve srovnání s výběry z velkých (resp. nekonečných) populací *ceteris paribus* menší výběr. Jak lze toto (na první pohled nelogické) pravidlo odůvodnit? Důvodem je skutečnost, že vzorce pro statistické testy, intervaly spolehlivosti či odhady velikosti výběrů vycházejí z předpokladu, že se provádí výběr s vrácením (vzorce jsou jednodušší), i když v praxi se běžně používají výběry bez vrácení. Vzorec rozptylu u výběru bez vrácení a výběru s vrácením se liší *konečnostním* násobitelem a právě tento faktor má v našem článku klíčovou roli. Jak vypadá vzorec konečnostního násobitele?

$$K = (N - n) / (N - 1), \quad (1)$$

kde  $N$  je velikost základního souboru a  $n$  velikost výběru.<sup>2</sup>

Za předpokladu, že vybíráme malou část populace (cca do 5–10 %), hodnota konečnostního násobitele se blíží jedné a lze jej zanedbat. Toto ovšem neplatí u výběrů z malých populací, kdy vybíráme podstatnou část základního souboru (platí to právě od cca 5–10 % vybraných jednotek) a kde již musíme konečnostní násobitel použít.

Konečnostní násobitel budeme v dalších částech využívat pro úpravu základních vzorců statistických testů či intervalů spolehlivosti, resp. budeme většinou užívat jeho druhé odmocniny, protože konečnostní násobitel se vztahuje k rozptylu a většina dále uvedených vzorců (pro testy a intervaly spolehlivosti) využívá standardní chybu, která je odmocninou rozptylu (technicky tedy směrodatnou odchylkou). Obecně lze tedy konstatovat, že velikost výběrového souboru z malé populace (s ohledem na požadovanou přesnost či sílu testů) se oproti výběrům z velkých populací může zmenšit, toto zmenšení udává právě hodnota založená na odmocnině konečnostního násobitele.

**Příklad 1.** Pro dvouvýběrový  $t$ -test s předpokládaným rozdílem mezi skupinami o velikosti cca poloviny směrodatné odchylky sledovaného ukazatele potřebujeme získat minimálně 150 výběrových jednotek, abychom rozdíl prokázali se silou testu minimálně 0,8 (srov. graf č. 3 a doprovodný text v Soukup & Kočvarová, 2016). Jak velký výběr bychom pro stejnou přesnost potřebovali, pokud víme, že naše populace má celkem 300 jednotek?

<sup>2</sup> Někdy se užívá i vzorce, který je mírným zjednodušením  $K = 1 - n / N$ . Pro většinu  $N$  a  $n$  dává tento vzorec v zásadě stejný výsledek.

- 128** Řešení: Dosadíme do výše uvedeného vzorce (1) pro  $K$  a získáme  $K = 0,502$ . Vypočteme-li odmocninu z  $K$ , získáme hodnotu  $0,708$ . Touto hodnotou násobíme velikost původního výběru ( $150$ ) a získáme hodnotu  $106$ , tj. v našem případě bude stačit jen výběr o velikosti  $106$  jednotek.

Poznámka: Taktó provedený výpočet je jen přibližný, více informací o dalších souvislostech čtenář nalezne v části věnované postupům pro dvouvýběrový  $t$ -test.

### 3 $t$ -testy (testy a intervaly spolehlivosti)

#### 3.1 Jednovýběrový $t$ -test

Na předchozí pojednání o konečnostním násobiteli navážeme situací, kdy chceme otestovat velikost průměru z jednoho výběru, potažmo pokud chceme počítat interval spolehlivosti pro průměr.<sup>3</sup> Klasický vzorec pro  $t$ -test je nutné analogicky předchozímu příkladu modifikovat dělením odmocninou z konečnostního násobitele, tj.

$$t_{kor}(I) = t / \sqrt{K}, \quad (2)$$

kde  $t$  je hodnota testového kritéria jednovýběrového testu a  $K$  je konečnostní násobitel. Pro výpočet je tedy potřebné znát výsledek  $t$ -testu spočtený běžným softwarem (např. SPSS či exceleem) a velikost populace, resp. našeho výběru.

**Příklad 2.** V našich datech získaných od 55 náhodně vybraných žáků chceme zjistit, zda průměrný počet hodin strávených před televizní obrazovkou jsou 3 hodiny. Je tato hypotéza udržitelná, pokud víme, že náš výběr byl získán z populace, která obsahuje jen 200 žáků, za předpokladu, že víme, že běžný výsledek  $t$ -testu má hodnotu  $t = 1,8$ , a je tedy statisticky neprůkazný ( $P = 0,078$ )?

Řešení: Dosadíme do výše uvedeného vzorce (2) pro  $t_{kor}(I)$ , v našem případě  $t = 1,8$ ;  $N = 200$ ;  $n = 55$ . Hodnota  $t_{kor}(I)$  je tedy  $2,1$  ( $P = 0,039$ ), znamená statisticky průkazný výsledek  $t$ -testu, tj. zamítneme hypotézu o tom, že žáci sledují televizi v průměru 3 hodiny. Pokud nechceme počítat výsledek ručně, stačí využít výstupů spočtených běžným softwarem (např. SPSS) a tyto dosadit do připravené pomůcky v excelu. Stačí zadat  $t$ ,  $N$  a  $n$  a vypočte se správně  $t_{kor}(I)$  a příslušná hodnota statistické významnosti (v buňce s označením Sig.).

Analogicky lze počítat interval spolehlivosti pro průměr v jedné skupině, musíme pouze znát velikost tohoto průměru a jeho směrodatnou odchylku v našich výběrových datech. Vzorec uvádí například Anděl (2003, s. 247):

$$x \pm t_{1-\alpha/2} \times s\sqrt{K} / \sqrt{n}, \quad (3)$$

<sup>3</sup> Analogický postup lze použít pro test či interval spolehlivosti proporce (např. podíl žáků s určitým názorem), je pouze třeba kódovat proměnnou, o kterou se zajímáme, 1 a 0 (dle toho, zda je přítomna námi sledovaná vlastnost). Poté lze využít vzorce a pomůcky dále uvedené.

kde  $\bar{x}$  je průměr hodnot v našem výběru,  $t_{1-\alpha/2}$  je  $(1-\alpha/2)$  procentní kvantil  $t$ -rozdělení,  $s$  směrodatná odchylka hodnot v našem výběru,  $K$  konečnostní násobitel dle vzorce (1) a  $n$  velikost našeho výběru.

**Příklad 3.** Pokračujme v příkladu 2 a určíme 95% interval spolehlivosti pro průměrný počet hodin strávených před TV obrazovkami, pokud víme, že průměr v našich výběrových datech činil 3,24 hodiny a směrodatná odchylka činila 0,99 hodiny.

Dosazením do vzorce (3) získáme interval spolehlivosti  $\langle 3,05; 3,43 \rangle$ , alternativně lze dosáhnout stejného výsledku dosazením do pomůcky v excelu. Meritorně tedy můžeme konstatovat, že průměrný počet hodin strávených před TV obrazovkou je o něco málo větší než 3 hodiny.

### 3.2 Dvouvýběrový $t$ -test

V případě dvouvýběrového  $t$ -testu není situace tak jednoduchá jako v případě jednovýběrového testu. Přibližně by šlo použít výše popsaných postupů (tj. vzorců 2, resp. 3) v případě, že srovnávané skupiny (výběry) by byly zhruba stejné velikosti, a dále za předpokladu, že by rozptyl v obou skupinách byl srovnatelný (z běžně dostupných testů lze ověřit zejména Leveneho testem). V ostatních případech je ale postup složitější a bude potřebné znát nikoli celkovou velikost zkoumané populace a našeho výběru, ale budeme muset znát tyto velikosti v obou srovnávaných skupinách. Nadto bude zapotřebí znát také směrodatné odchylky v obou výběrech a průměry v obou našich srovnávaných skupinách. Vzorce zde pro jejich komplikovanost již nebudeme uvádět a odkážeme jen na literaturu (např. Hendl, 2012, s. 221) a vše vypočteme přímo v připravené pomůcce v excelu.

**Příklad 4.** V návaznosti na příklady 2 a 3 zkusíme zjistit, zda je možné, že dívky a chlapci sledují televizní obrazovku v průměru stejně dlouho. Pro jednoduchost předpokládejme, že v naší populaci je shodně chlapců a dívek ( $N_1 = N_2 = 100$ ) a v našem výběru máme  $n_1 = 30$  chlapců a  $n_2 = 25$  dívek. Průměrná doba strávená před TV pro chlapce činila v našem výběru 3,47 hodiny a pro dívky 2,96 hodiny, směrodatná odchylka pro chlapce 1,1 a pro dívky 0,85 hodiny. Kromě testu vypočteme i interval spolehlivosti pro rozdíl mezi průměry (obojí umí připravená pomůcka v excelu).

Po dosazení hodnot do pomůcky v excelu vyjde hodnota testového kritéria  $t_{kor}(II) = 2,27$ , výsledek testu je tedy na 5% hladině statisticky významný ( $P = 0,026$ ). Interval spolehlivosti pro rozdíl průměrů je  $\langle 0,06; 0,96 \rangle$ .

Dodejme, že analogický postup lze použít pro test či interval spolehlivosti rozdílu dvou proporcí (např. podíl chlapců a dívek s určitým názorem – viz poznámka 3). Pro případ párového  $t$ -testu se uplatní stejný postup jako u jednovýběrového  $t$ -testu, není zde tedy třeba dalšího popisu ani vzorců či speciálních pomůcek, stačí využít výše uvedené (tj. vzorce 2 a 3 a postupy v příkladech 2 a 3).

## 4 Analýza rozptylu (základní test)

Poslední procedurou sloužící k vyhodnocení průměrů, která zde bude pojednána, je analýza rozptylu. S ohledem na komplexnost této techniky bude pozornost zaměřena jen na základní test skrze analýzu rozptylu pro případ jediného faktoru (stranou bude ponechána vícefaktorová analýza rozptylu, vícerozměrná analýza rozptylu či analýza rozptylu pro opakovaná měření). I v případě jednofaktorové analýzy rozptylu bude ukázána jen korekce pro základní test (nadto pro poměrně specifický případ) a u mnohonásobných srovnání (tzv. post hoc comparisons) bude odkázáno na poznatky o dvouvýběrovém testu. Pokud používáme analýzu rozptylu pro výběry, které jsou velkou částí populace, a platí (alespoň přibližně), že v jednotlivých skupinách je obdobně pozorování, lze užít velice jednoduchou korekci testového kritéria  $F$  dle vzorce

$$F_{kor} = F / K, \quad (4)$$

kde  $F$  je testové kritérium základního testu v analýze rozptylu (vypočtené např. pomocí SPSS či excelu) a  $K$  je konečnostní násobitel dle vzorce (1).

Pokud tedy známe velikost populace  $N$ , velikost našeho výběru  $n$  a  $F$ , lze vypočítat  $F_{kor}$  a pomocí statistických tabulek jej vyhodnotit. Alternativně (a snadněji) je možné využít pomůcku v excelu. Vše opět ilustrujeme na praktickém příkladu.

**Příklad 5.** Naše populace má 300 jednotek ( $N = 300$ ) a náš výběr obsahuje 3 skupiny po 50 jednotkách ( $n = 150$ ). Porovnáme-li průměry těchto 3 skupin  $F$ -testem v analýze rozptylu, získáme výsledek  $F = 1,7$ . Pro vyhodnocení na 5% hladině statistické významnosti je potřebný 95% kvantil  $F$  rozdělení  $F_{0,95}(2, 147)^4 = 3,05$ , a tudíž bychom konstatovali, že výsledek analýzy rozptylu je statisticky nevýznamný, tj. nelze zamítnout (nulovou) hypotézu o shodě průměrů ve skupinách (v námi sledované populaci). Pokud zohledníme, že naše populace je malá ( $N = 300$ ) a výběr z ní velký, a použijeme  $F_{kor}$ , pak je hodnota  $F_{kor} = 3,39$  a tato hodnota již svědčí ve prospěch přijetí alternativní hypotézy, tj. konstatujeme, že minimálně dva průměry jsou statisticky významně odlišné (zda jsou odlišné věcně, poznáme orientačně z jejich velikostí). Excelovská pomůcka nám nadto pomůže vypočítat  $P$  hodnotu pro  $F_{kor}$ , její hodnota činí 0,036.

Pokud v analýze rozptylu dospějeme k zamítnutí nulové hypotézy, pokračujeme zpravidla následnými testy. Zde je situace velice složitá (testů existují desítky, dělí se minimálně na testy pro případ shodných a neshodných rozptylů) a nelze ji postihnout na ploše jednoho článku. Pro orientační výpočty statistické významnosti rozdílů mezi jednotlivými dvojicemi pak lze využít dvouvýběrové  $t$ -testy (viz popis výše) s tzv. Bonferroniho korekcí (srov. Hendl, 2012, s. 205), která je založena na skutečnosti, že jednotlivé  $t$ -testy vyhodnocujeme při snížené hladině statistické

<sup>4</sup> V závorce jsou počty stupňů volnosti, první hodnota odpovídá počtu skupin zmenšenému o jednotku ( $2 = 3 - 1$ ), druhá velikosti výběru zmenšenému o počet skupin ( $147 = 150 - 3$ ).

významnosti (ne tedy např. běžně užívanou hodnotou 0,05). Toto snížení provedeme tak, že námi užívanou hladinu statistické významnosti dělíme počtem prováděných srovnání (při  $k$  skupinách je počet srovnání dán vzorcem  $k \times (k - 1) / 2$ ). Tuto část již ponecháme bez příkladu.

## 5 Korelace (test a interval spolehlivosti pro korelační koeficient)

V případě, že užíváme Pearsonův korelační koeficient (tj. nejčastěji užívaný koeficient pro měření lineární souvislosti) a máme výběr velké části malé populace, musíme obdobně jako v předchozích případech uplatnit konečnostní násobitel. Postup se ovšem liší u testu statistické významnosti korelačního koeficientu (typicky vycházející z nulové hypotézy o nesouvislosti, resp. o nulové hodnotě korelace v populaci) a u výpočtu intervalu spolehlivosti (čtenář, který není seznámen s intervaly spolehlivosti u korelačních koeficientů, si může více přečíst např. v učebnici Hendla, 2012, s. 262–263, nebo v článku Soukupa, 2010).

Začněme nejdříve testem o nulové hodnotě korelačního koeficientu. Testové kritérium se zde vypočte dle vzorce

$$t = [r \times \sqrt{n - 2}] / \sqrt{1 - r^2}, \quad (5)$$

kde  $r$  je hodnota korelačního koeficientu v našem výběru a  $n$  velikost našeho výběru.

Pokud je náš výběr velkou částí malé populace, musíme vzorec modifikovat odmocninou konečnostního násobitele ( $\sqrt{K}$ ) ve jmenovateli následovně:

$$t_{kor} (III) = [r \times \sqrt{n - 2}] / [\sqrt{1 - r^2} \times \sqrt{K}], \text{ resp. } t_{kor} (III) = t / \sqrt{K}, \quad (6)$$

kde  $K$  je konečnostní násobitel (vzorec 1) a ostatní symboly mají totožný význam jako ve vzorci (5).

Poté musíme vyhodnotit hodnotu  $t_{kor} (III)$ , u výběrů do velikosti cca 30 kvantilem  $t$  rozdělení s počtem stupňů volnosti o 2 jednotky nižším, než je velikost výběru ( $n - 2$ ), u výběrů nad 30 jednotek lze pak použít přibližně kvantilů standardizovaného normálního rozdělení, tj. nejčastěji pro testování na 5% hladině oboustranným testem hodnotu 2 (či přesněji 1,96). Nejsnazší bude opět použít připravenou pomůcku v excelu.

**Příklad 6.** V našich datech získaných od 55 náhodně vybraných žáků jsme spočítali korelaci mezi dvěma postoji o hodnotě 0,25. Je tato korelace statisticky průkazná, pokud víme, že náš výběr byl získán z populace, která obsahuje jen 200 žáků?

Řešení: Dosadíme do výše uvedeného vzorce (5) pro  $t$ , v našem případě  $r = 0,25$ ,  $N = 200$ ,  $n = 55$ . Hodnota  $t$  je 1,879, což by znamenalo, že korelace je statisticky neprůkazná. Pokud ovšem spočteme korektně  $t_{kor} (III) = 2,2$ , jde již o statisticky průkazný výsledek. Pokud nechceme počítat každý výsledek ručně, stačí využít výstupů

132 spočtených v běžném softwaru (např. SPSS) a tyto dosadit do připravené pomůcky v excelu. Stačí zadat  $N$ ,  $n$  a  $r$ ; nebo  $N$ ,  $n$  a  $t$ , poté se vypočte správně  $t_{kor}$  (III) a příslušná hodnota statistické významnosti (v buňce s označením Sig.).

Poznámka: Korelace o hodnotě 0,25 nebude v pedagogickém výzkumu většinou zcela bez věcného významu, nicméně nejde o hodnotu převratně vysokou. Nezapomeňme, že kromě interpretace statistické významnosti je stejně tak potřebné se věnovat interpretaci věcné významnosti výsledku.

U korekce pro výpočet intervalu spolehlivosti je situace složitější než v případě výpočtu testu o nulové hodnotě korelace. Interval spolehlivosti nelze (běžně) spočítat přímo, ale užívá se nejružnějších transformací (nejčastěji tzv. Fisherovy, podrobnosti viz Soukup, 2010). Právě na tyto transformace je nutné uplatnit korekci s využitím konečnostního násobitele, resp. jeho odmocniny. Nebudeme dále zabíhat do výpočetních podrobností (čtenář znalý středoškolské matematiky a funkcí v excelu si vše může dovést z pomůcky v excelu). Ilustrujeme princip opět na jednoduchém příkladu.

**Příklad 7.** Naším úkolem je spočítat 95% interval spolehlivosti pro zadání z minulého příkladu. Pokud bychom nezohledňovali konečnostní násobitel, získali bychom následující interval:  $(-0,03; 0,49)$ , srov. například výpočet on-line kalkulátorem na adrese: <http://www.how2stats.net/2011/09/confidence-intervals-for-correlations.html>. Při zohlednění skutečnosti, že naše data pocházejí z malé populace, bude výsledný interval užší a nebude obsahovat nulu:  $(0,02; 0,45)$ , viz pomůcka v excelu.

Poznámka: Věcně nás jak nekorigovaný, tak i korigovaný interval spolehlivosti v našem příkladu vedou k výrazné opatrnosti, protože máme poměrně malý datový soubor, a naše zjištění je tedy krajně nespolehlivé.

Dodejme pro úplnost, že popsané postupy a pomůcky lze uplatnit i pro korelace ordinálních veličin počítané dle Spearmanova vzorce. Pro jiné koeficienty však takto přímé úpravy neplatí a popsané postupy či pomůcku v excelu zde přímo využít nelze.

## 6 Regrese (test a interval spolehlivosti pro regresní koeficient)

Obdobně jako v případě korelačních koeficientů je možné využít konečnostní násobitel ke korekcím standardních chyb pro výpočty statistické významnosti u regresních koeficientů, resp. pro výpočty jejich intervalů spolehlivosti. Situace je zde o poznání snazší než u korelací, protože zde není zapotřebí žádných transformací a stačí jen upravit standardní chyby odhadu a poté dopočítat statistickou významnost či interval spolehlivosti. Úprava standardní chyby odhadu je snadná, čtenář jistě sám z předchozích ukázek nahlédne, že ji vynásobíme odmocninou konečnostního násobitele. Pokud tedy spočítáme klasicky tabulku s regresními koeficienty (např. v SPSS či excelu), zkopírujeme hodnoty koeficientů (většinou označené B) a jejich standardních chyb (nejčastěji označených S.E.), poté je využijeme pro další ruční výpočty, případně pro výpočty v pomůcce v excelu.



**Příklad 8.** V našich datech získaných od 100 náhodně vybraných žáků jsme spočítali regresní model s dvěma nezávisle proměnnými, výsledek zobrazuje následující tabulka 1. Jsou vlivy nezávisle proměnných statisticky průkazné? Jaké jsou intervaly spolehlivosti těchto regresních koeficientů, pokud víme, že náš výběr byl získán z populace, která obsahuje jen 200 žáků?

Zadání obsahuje tabulka 1 v levé části, v pravé části je pak výpočet korigované standardní chyby a v další tabulce 2 jsou následné výsledky statistických testů, resp. intervalů spolehlivosti.

**Tabulka 1** Fiktivní výsledek lineární regresní analýzy s konstantou a dvěma proměnnými

Proměnná	B	S.E.	Korigované S.E.
Konstanta	500	50	
Proměnná 1	140	80	$= 80 \times [(200 - 100) / (200 - 1)] = 56,710$
Proměnná 2	25	15	$= 15 \times [(200 - 100) / (200 - 1)] = 10,633$

Pozn.: Poslední sloupec (Korigované S.E.) obsahuje výpočty, nikoli zadání.

Testy uvedené v tabulce 2 (označované jako  $t$ , neboť testové kritérium má  $t$  rozdělení) pro jednotlivé regresní koeficienty získáme tak, že vydělíme hodnotu  $B$  korigovanou hodnotou  $S.E.$  Tyto hodnoty pak porovnáme s kvantily  $t$ -rozdělení<sup>5</sup> (pro výběr větší než 30 pak standardizovaného normálního rozdělení) a rozhodneme o výsledku statistického testu. Interval spolehlivosti (95 %) poté spočteme dle vzorce  $B \pm 1,96 \times (\text{Korigované } S.E.)$ . Takto vypočtené výsledky pro naše data (získané opět skrze excelovskou pomůcku) jsou zobrazeny v tabulce 2.

**Tabulka 2** Výpočty statistického zhodnocení vlivu proměnných v lineární regresní analýze s konstantou a dvěma proměnnými

Proměnná	T	Sig.	Interval spolehlivosti	
			dolní mez	horní mez
Konstanta				
Proměnná 1	2,469	0,015	27,445	252,555
Proměnná 2	2,351	0,021	3,896	46,104

Pozn.: Sloupce kurzivou obsahují výsledky výpočtů.

Dodejme, že po korekci odmocninou konečnostního násobitele je na 5% hladině statistické významnosti vliv obou proměnných průkazný, před touto korekcí tomu však bylo naopak (není zobrazeno v tabulce 1, pečlivý čtenář však snadno dopočte sám).

Výsledná tabulka určená k publikaci tedy bude sloučením dvou předchozích tabulek a bude mít podobu uvedenou v tabulce 3.

<sup>5</sup> Pro jeho využití musíme znát počet odhadovaných parametrů, tj. počet řádků v tabulce s regresními koeficienty, běžně označovaný jako  $p$ . Tento parametr je vyžadován v pomůcce v excelu jako povinný.

## 134 Tabulka 3 Ukázka publikačního výstupu v lineární regresní analýze

Proměnná	B	S.E. korig.	T	Sig.	Interval spolehlivosti	
					dolní mez	horní mez
Konstanta	500					
Proměnná 1	2,469	56,710	2,469	0,015	27,445	252,555
Proměnná 2	2,351	10,633	2,351	0,021	3,896	46,104

Opět je namístě poznámka o velké šíři vypočtených intervalů spolehlivosti, důvodem je zejména skutečnost, že náš výběr je poměrně malý. Dodejme, že technicky je možné korigovat i standardní chybu pro konstantu a počítat následně test a interval spolehlivosti. Nicméně ve většině aplikací se s konstantou nepracuje a meritorně se neinterpretuje, proto zde i v excelovské pomůcce výpočty chybí (nicméně čtenář znalý excelu velice snadno zkopíruje vzorce a výpočty by se pak provedly i pro konstantu).

## 7 Kontingenční tabulky a chí-kvadrát test nezávislosti

V případě užívání kontingenčních tabulek nejčastěji řešíme problém souvislosti dvou nominálních či ordinálních proměnných, k jejímu posouzení využíváme chí-kvadrát test o nezávislosti znaků. Pokud náš výběr pochází z malé populace, pak je opět zapotřebí výsledky tohoto testu korigovat o konečnostní násobitel a takto korigované testové kritérium poté vyhodnotit oproti kvantilům  $\chi^2$  rozdělení (počet stupňů volnosti je dán násobkem o jednotku zmenšenými počty řádků a sloupců). S ohledem na charakter testového kritéria  $\chi^2$  (statisticky jde o součet druhých mocnin normálně rozdělených veličin, technicky v případě tohoto testu sčítáme rezidua v kontingenční tabulce) musíme korigovat testové kritérium nikoliv odmocninou konečnostního násobitele, ale přímo konečnostním násobitelem (srov. např. Rao & Thomas, 1989). Vzorec pro korigované testové kritérium (označme jej analogicky k předchozím  $\chi^2_{kor}$ ) bude následující:

$$\chi^2_{kor} = \chi^2 / K, \quad (7)$$

kde  $\chi^2$  je hodnota testového kritéria spočtená klasickým způsobem a  $K$  je konečnostní násobitel (viz vzorec 1).

Pro ilustraci opět uvedeme jednoduchý příklad.

**Příklad 9.** Při výzkumu specifické populace žáků čítající 200 jednotek byl proveden náhodný výběr 100 žáků. Od těchto 100 žáků jsme získali odpovědi. Jeden z výsledků (týkající se názoru na zvýšení počtu hodin matematiky) našeho výzkumu prezentuje následující tabulka 4.

**Tabulka 4** Fiktivní kontingenční tabulka pro měření souvislosti pohlaví a postojů k navýšení počtu hodin matematiky (absolutní četnosti)

	Souhlas	Nesouhlas	Celkem
Dívky	21	24	45
Hoši	35	20	55
Celkem	56	44	100

První náhled na tabulku jasně říká, že chlapci spíše se zvýšením počtu hodin matematiky souhlasí, dívky naopak spíše nesouhlasí, a rozdíl je poměrně výrazný. Nicméně pro zobecnění výsledků z výběru na populaci je třeba mj. užít statistický test. Klasický výpočet testového kritéria poskytne hodnotu  $\chi^2 = 2,89$ , což je při jednom stupni volnosti na 5% hladině významnosti *statisticky nevýznamný* výsledek ( $P = 0,089$ ). Meritorně bychom tedy konstatovali, že rozdíl v názorech mezi chlapci a dívkami nebyl statisticky prokázán. Nicméně my víme, že náš výběr pochází z malé populace, a měli bychom výpočty upravit. Při korekci konečnostním násobitelem dle vzorce (7) je  $\chi^2_{kor} = 5,76$ , což je při jednom stupni volnosti na 5% hladině významnosti *statisticky významný* výsledek ( $P = 0,016$ ). Dodejme, že dopad je díky použití konečnostního násobitele (ne tedy jeho odmocniny) v případě kontingenčních tabulek výraznější. Pro opakované výpočty je opět výhodné použít připravenou pomůcku v excelu, kam je třeba zadat hodnotu spočteného  $\chi^2$  kritéria (např. výsledek z SPSS či excelu), velikost populace a výběru ( $N$ , resp.  $n$ ) a dále počet řádek ( $k$ ) a počet sloupců ( $l$ ) kontingenční tabulky. Pomůcka sama vypočte  $\chi^2_{kor}$  a  $P$  hodnotu.

## 8 Další software pro práci s výběry z malých populací

Samozřejmě, že výše uvedené vzorce a postupy nepokrývají všechny užívané statistické techniky. Nadto práce s ručními výpočty či excelelem nemusí být pro každého přijatelná. Pak nezbyvá než užít speciální software či různé doplňky ke statistickému softwaru. Zmíníme zde základní možnosti.

Speciální software, který mj. umí korektně přímo pracovat s daty vybíranými z malých populací, je SUDAAN.<sup>6</sup> Tento program buď funguje samostatně, nebo existuje jako doplněk k softwaru SAS. Jde o placený software (kupuje se každoroční licence), jehož velkou předností je schopnost fungovat s nejrůznějšími výběrovými designy, pro pedagogické výzkumníky může být vhodné, že umí mj. načítat data ve formátu SPSS.

Další možností je využití SAS či STATA, kde jsou některé postupy též implementovány, nicméně jde o poměrně drahé programy a uživatel se musí naučit s nimi pracovat. Poslední možností (ale nikoli z hlediska důležitosti) je využít balíčky připravené v prostředí R. R je stejně jako jednotlivé balíčky zcela zdarma a díky nadšení jednotlivých uživatelů lze nalézt na internetu mnohé rady. Pro popisované situace se

<sup>6</sup> Zájemce se může více dozvědět na webu: <https://www.rti.org/impact/sudaan-statistical-software-analyzing-correlated-data>.

- 136 hodí zejména balíček *survey*.<sup>7</sup> Komplikací používání R může být poměrně specifický příkazový jazyk a také omezený formát dat, který načteme (nicméně se speciálním balíčkem lze načítat i SPSS data). Ze všech představených nástrojů je R s balíčky při práci s výběry z malých populací nejuvšestrannější.

## 9 Shrnutí a omezení

Text si kladl za cíl položit teoretické a praktické základy pro korektní statistickou práci s daty, která byla získána z malých populací, pokud z této populace vybíráme nezanedbatelnou část (nad 5 %). Samozřejmě by bylo žádoucí rozšířit okruh statistických technik i na další, nicméně tím by text výrazně vykročil z objemu určeného pro běžné články. Nadto není cílem připravit pro všechny situace speciální pomůcky, sofistikovanější výzkumník se může nechat inspirovat a sáhnout po softwaru, který je doporučován v předchozí části, a může tak počítat se složitějšími statistickými technikami (např. logistickou regresí). Další vhodnou strategií je též konzultovat složitější statistické problémy se statistiky, kteří by měli umět pomoci.<sup>8</sup> Zcela jistě nelze od pedagogického výzkumníka žádat, aby uměl vše, nadto ještě ve všech možných programech.

### Literatura

- Anděl, J. (2003). *Statistické metody*. Praha: MatfyzPress.
- Hendl, J. (2012). *Přehled statistických metod: analýza a metaanalýza dat*. Praha: Portál.
- Rao, J. N. K., & Thomas, D. R. (1989). Chi-squared tests for contingency tables. In C. J. Skinner, D. Holt, & T. M. F. Smith (Eds.), *Analysis of complex surveys* (s. 89–114). Chichester: Wiley.
- Soukup, P. (2010). Nesprávná užívání statistické významnosti a jejich možná řešení. *Data a výzkum – SDA Info*, 4(2), 77–104.
- Soukup, P., & Kočvarová, I. (2016). Velikost a reprezentativita výběrového souboru v kvantitativně orientovaném pedagogickém výzkumu. *Pedagogická orientace*, 26(3), 512–536.

Petr Soukup  
petr.soukup@fsv.cuni.cz

<sup>7</sup> Detailní popis nalezne zájemce na webových stránkách: <https://cran.r-project.org/web/packages/survey/survey.pdf>.

<sup>8</sup> Samozřejmě se lze obrátit i na autora článku.